A study of correlations in the stock market

Chandradew Sharma, Kinjal Banerjee *

Department of Physics, BITS Pilani K.K Birla Goa Campus, N.H. 17B Zuarinagar, Goa 403726, India

HIGHLIGHTS

- Study of BSE returns over 8 years.
- Distribution of returns is not normal.
- Persistence of ACF.
- Large cross correlation due to external news.

ARTICLE INFO

ABSTRACT

We study the various sectors of the Bombay Stock Exchange (BSE) for a period of 8 years from April 2006 to March 2014. Using the data of daily returns of a period of eight years we make a direct model free analysis of the pattern of the sectorial indices movement and the correlations among them. Our analysis shows significant auto correlation among the individual sectors and also strong cross-correlation among sectors. We also find that auto correlations in some of the sectors persist in time. This is a very significant result and has not been reported so far in Indian context. These findings will be very useful in model building for prediction of price movement of equities, derivatives and portfolio management. We show that the Random Walk Hypothesis is not applicable in modeling the Indian market and mean–variance–skewness–kurtosis based portfolio optimization might be required. We also find that almost all sectors are highly correlated during large fluctuation periods and have only moderate correlation during normal periods.

1. Introduction

The stock market is an extremely complex system with various interacting components [1]. The movement of stock prices are somewhat interdependent as well as dependent on a wide multitude of external stimuli like announcement of government policies, change in interest rates, changes in political scenario, announcement of quarterly results by the listed companies and many others. The overall result is a chaotic complex system which has so far proved very difficult to analyze and predict. In fact it is still not completely clear, what are the generic features that will appear in any stock market and what are the features which depend on the social, political and economic climate of the country and/or of the world. So it is important to study each market individually so that finally we can be sure that certain behaviors or patterns are universal. Although some amount of work has been done in understanding the stock markets in Europe [2] and the United States [3], the proper mathematical and statistical study of emerging markets like India are in their infancy [4].

So far, there is no exact understanding on which external stimulus has how much effect on the stock prices or even how the self interactions of the various stocks or the various sectors drive the market. Broadly speaking, the price movement of a

---

particular stock can be classified as (i) market (common to all stocks), (ii) sector (related to a particular business sector) and (iii) idiosyncratic (limited to an individual stock). While it is virtually impossible to develop any theory for this idiosyncratic movement, it is possible to analyze, study and build models for the other two types of stock movement. From an investors point of view, the most important reason to understand the stock market is to get the maximum possible return on an investment with the minimum possible risk. So a better understanding of the stock market will lead to better theories of portfolio management.

One important step in improving our understanding of the stock market is to study how the stock price movement of one stock affects the price of other stocks. One way to do this would be to see how one stock movement affects the others within the same sector. Another is to study how the overall prices of the various sectors are correlated. The goal of this study is to try to determine and quantify, from the available data, some of the possible correlations which might exist between the stock prices. This will not only enhance the understanding of the stock market as a whole but will play a crucial role in investment decisions like portfolio management. A systematic model independent analysis of the data that we do will also help in building more efficient and enhanced models which will give adequate weightage to the various relations which exist between movement of stock prices across sectors in a market and may help in forecasting future trends. Studies of such correlations have been carried out to a limited extent in the context of New York Stock Exchange [5], but to the best of our knowledge, no such study exists for the Bombay Stock Exchange (BSE) [6].

To understand the financial market, it is very important to know the distribution of the return on a stock. Our data consists of the daily returns of 12 sectors of stocks of the BSE for Financial Year (FY) 2006 to FY 2013 i.e. 1990 days from 3rd April 2006 to 31st March 2014. We will be treating each sector as one entity in the rest of the paper. This approach is novel and has not been carried out before, at least in the context of Indian markets.

If $P_i(t)$ is the index of the sector $i = 1, \ldots, N$ at time t, then the *(logarithmic) return* of the ith sector over a time interval $t = 1$ to $t = T$ days in the interval is defined as

$$R_i(t) \equiv \ln P_i(t + 1) - \ln P_i(t). \tag{1}$$

In our case $T = 1900$, the number of days we have considered, and $N = 13$ because we look at the following 12 sectors S&P BSE Auto (Auto), S&P BSE Bankex (Bankex), S&P BSE Consumer Durables (CD), S&P BSE Capital Goods (CG), S&P BSE FMCG (FMCG), S&P BSE Health care (HC), S&P BSE IT (IT), S&P BSE Metal (Metal), S&P BSE Oil and Gas (Oil and Gas), S&P BSE Power (Power), S&P BSE Realty (Realty) and S&P BSE Teck (Teck) and the S&P BSE SENSEX (Sensex) which serves as the benchmark. The plot of the Sensex index and the log return over the time interval under consideration is given in Fig. 1. From Fig. 1, it is clear that we can divide the entire period in two sub interval (i) from FY 2006 to FY 2009 as large fluctuation period and (ii) from FY 2010 to FY 2013 for normal period. We shall discuss how the cross correlations of the sectors are markedly different in these two periods, later in the paper.

Obviously the mean return of the ith sector is given by

$$\bar{R}_i = \frac{1}{T} \sum_{t=1}^{T} R_i(t). \tag{2}$$

Defining $R_i' = (R_i(t) - \bar{R}_i)$, we can write the kth moment of the ith sector as

$$m_k(i) = \frac{1}{T} \sum_{t=1}^{T} (R_i'(t))^k. \tag{3}$$

For example, the second moment gives the variance as

$$\sigma(i) = \frac{1}{T} \sum_{t=1}^{T} (R_i'(t))^2. \tag{4}$$

These definitions are used in the analysis subsequently.

Our paper is organized as follows. In Section 2 we explore the individual sectors mentioned above and use the data to determine some features of the distribution of the returns and find significant deviations from normality. We then calculate the auto correlation of log returns for all sectors indices, to test the market efficiency, and find that there is significant auto correlation in most of the sectors of BSE at lag 1. The more surprising result is that the analysis of our data shows that the auto correlations in some sectors persist at higher lags. In Section 3 we analyze the cross-correlations among sectors in BSE. Our study spans over FY 2006–2013, a time span which consisted a period large fluctuation in indices movement and normal fluctuation period. We find that, almost all sectors are highly correlated during period 2006–2009 and they are moderately correlated during 2009–2013. We finally conclude in Section 4 with a summary of our results and its interpretations.

2. Understanding BSE sectors

It is commonly believed that the distribution for log return of a stock or for log change in a index movement is a normal distribution. However, many empirical studies shows deviation from this perception. Consequently, any prediction based on the normal distribution will generally fail. In particular, if there is any deviation from normality, the Random Walk
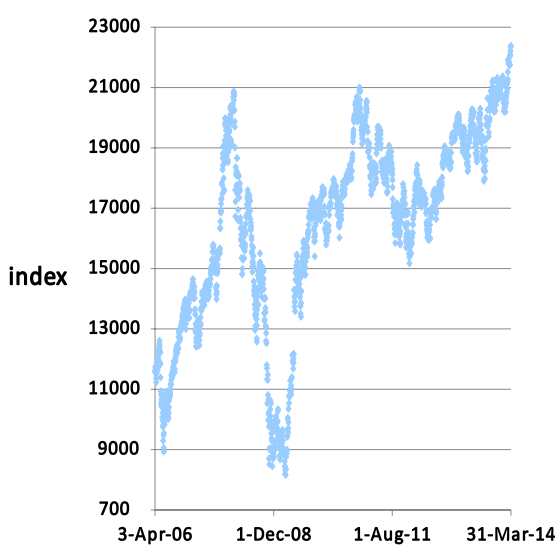
(a) Variation of BSE index over time.
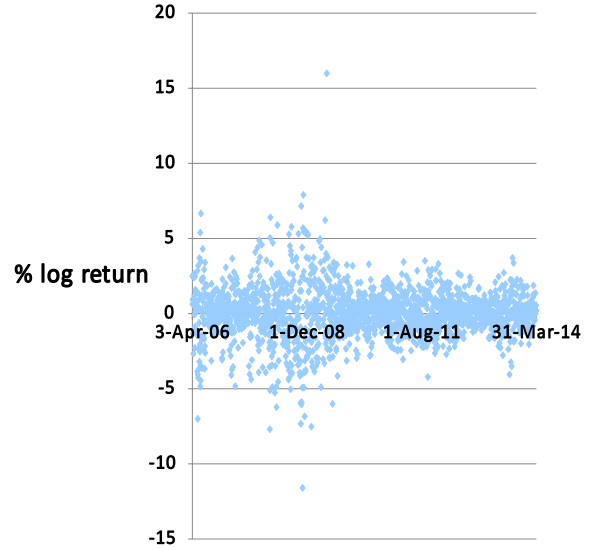
(b) Variation of (log) return of BSE over time.
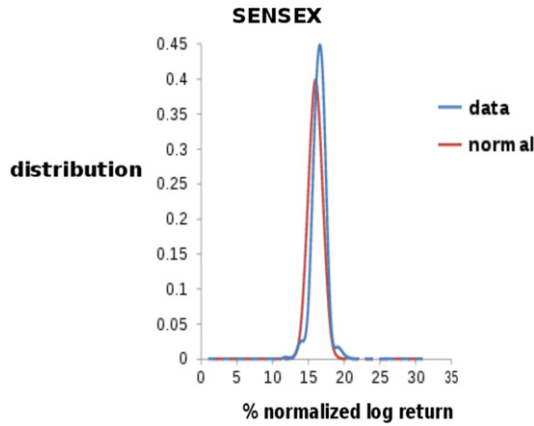
(c) Variation of (log) return of BSE over time.

Fig. 1. Behavior of S&P BSE SENSEX for April 2006–Mar 2014.

Hypothesis will not be valid. Therefore, it is essential to first understand the distribution of any stock or index movement before using any model. Let us first consider the distribution of returns for the various sectors.

The study of skewness and kurtosis is very useful to characterize the distribution. We know that if a distribution is normal, then sample skewness and sample excess kurtosis will be close to zero [7]. Any significant deviation from zero indicates a deviation from normality.

The sample skewness and excess kurtosis of any distribution of the ith sector can be written in terms of the moments (3) as

$$\text{Sample skewness: } G_1(i) = \frac{T(T-1)}{(T-2)} \frac{m_3(i)}{m_2(i)^{3/2}} \tag{5}$$

$$\text{Sample excess Kurtosis: } G_2(i) = \frac{(T-1)}{(T-2)(T-3)} \left[(T+1) \left(\frac{m_4(i)}{m_2(i)^2} - 3 \right) + 6 \right]. \tag{6}$$

The Standard Error in Skewness (SES) and Standard Error in Kurtosis (SEK) are given by [7]

$$\text{SES} = \sqrt{\frac{6T(T-1)}{(T-2)(T+1)(T+3)}}; \qquad \text{SEK} = 2\text{SES}\sqrt{\frac{T^2-1}{(T-3)(T+5)}}. \tag{7}$$
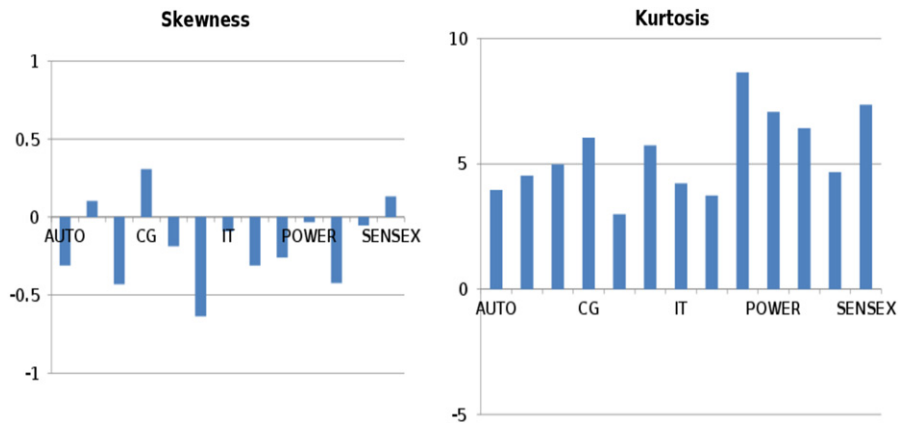
Fig. 2. Sample skewness and sample excess kurtosis of all the sectors for April 2006–Mar 2014.

The sample skewness and sample excess kurtosis of the above sectors are displayed in Fig. 2. The Standard Error in Skewness (SES = 0.06) and Standard Error in Kurtosis (SEK = 0.11) are calculated based on the formulae given by (7). It is clear from Fig. 2 that there is significant deviation from zero for sample skewness and sample excess kurtosis in all sectors. Hence, based on the study of sample Skewness and sample excess kurtosis, we can say confidently that each individual sector's return shows positive kurtosis (fat tails) accompanied by skewness. This clearly shows that the returns of none of the sectors are normally distributed.

To further strengthen this claim we perform the D'Agostino–Pearson omnibus test [7]. It is based on two quantities depending on both skewness and kurtosis. The quantities are defined as follows:

$$Z_{G1}(i) = \frac{G1}{\text{SES}}; \qquad Z_{G2}(i) = \frac{G2}{\text{SEK}} \tag{8}$$

$$DP(i) = (Z_{G1})^2 + (Z_{G2})^2. \tag{9}$$

If the distribution of ith sector is normal, the $DP(i)$ should be that of χ distribution. So if $DP(i) > \chi^2_{critical}$, then the distribution of the ith sector is not normal. For a normal distribution of ith sector, $\chi^2_{critical}(2df)$ should be 13.82 with significance level of 0.1%. What we find is that the DP values for all the sectors are much larger than 13.82. Therefore, the statistical results clearly indicate that the data does not satisfy the normality assumption, i.e. the change in index movement of individual sectors shows large deviation from normal distribution. This finding is also consistent with recent works [8] and shows that returns are driven by asymmetric and fat-tailed distribution. This also clearly indicates that the market cannot be modeled using the Random Walk Hypothesis [9]. For stock market modeling or from the perspective of portfolio management the mean–variance model [10] should be expanded by mean–variance–skewness–kurtosis based portfolio optimization [11].

To further explore the nature of the auto correlation, we look at the time series of the auto correlation data for the various sectors. This study is important because if there is auto correlation in the time series we can predict immediate future based on present information. If there is no auto correlation in the time series data, the data are uncorrelated and it is not possible to make future predictions confidently.

To emphasize, if there is auto correlation in the time series at lag 1, it is possible to make predictions about immediate future with high degree of certainty. Here, we have estimated the sample auto covariance at lag k for a finite ith time series $R_i(t)$ of T observations by [12]

$$\gamma_{ik} = \frac{1}{T} \sum_{t=1}^{T-k} (R_i(t) - \bar{R}_i)(R_i(t+k) - \bar{R}_i) \tag{10}$$

where $R_i(t)$ is given by our definition (1). The auto correlation at lag k can then be estimated as:

$$\rho_{ik} = \frac{\gamma_{ik}}{\gamma_{i0}}. \tag{11}$$

The function ρ_{ik} is known as the Auto Correlation Function (ACF).

We have used the Bartlett's approximation [13] to estimate the variance of the ACF, at lags k greater than some value q beyond which the auto correlation function may be deemed to have *died out*. This is defined as [12]:

$$\text{var}[\rho_{ik}] \approx \frac{1}{T} \sum_{v=1}^{q} (1 + 2\rho_{iv}^2) \quad k > q. \tag{12}$$

(a) ACF for FMCG.

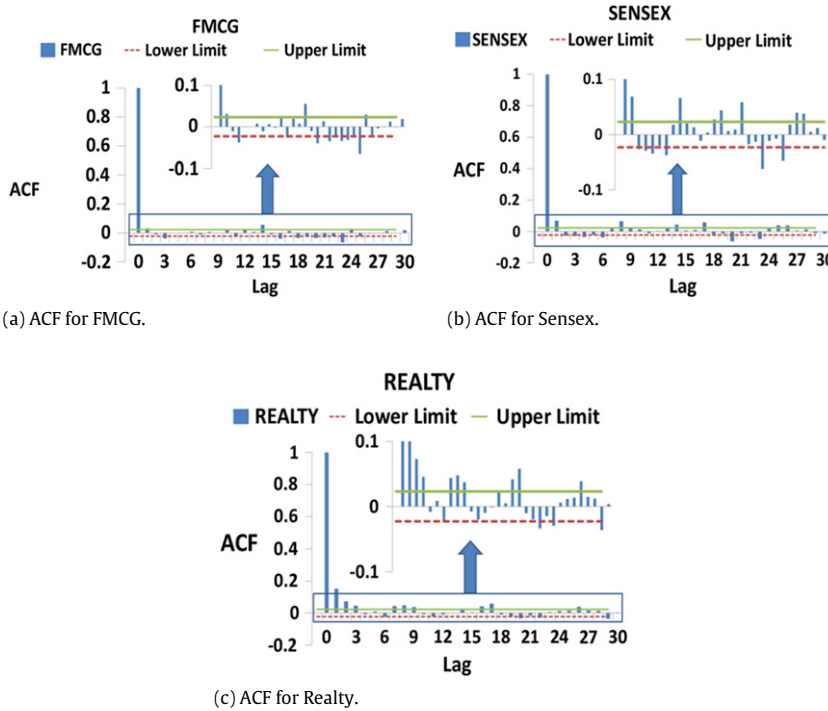(b) ACF for Sensex.

(c) ACF for Realty.

Fig. 3. ACF for FMCG, BSE SENSEX and REALTY sectors for April 2006–Mar 2014.

The standard error for estimated auto correlation ρ_{ik} is:

$$SE[\rho_{ik}] = \sqrt{\text{var}[\rho_{ik}]}. \tag{13}$$

We calculate the auto correlation of log returns for all sector indices of BSE. It clearly shows that there is significant auto correlation in most of the sectors of BSE at lag 1. Therefore, residual effect is confirmed in almost all sectors in BSE. A statistically significant ACF value at lag 1 indicates an autoregressive component exists in the time series. In fact, we find some auto correlation in most of the sector persists over time.

Our results show that there is very little auto correlation in FMCG, weak auto correlation in IT, Teck and Oil and Gas, and significant lag 1 auto correlation in Auto, Bankex, CD, CG, HC, Metal, Power, Realty, and also in Sensex. In Fig. 3, we have plotted the ACF for three BSE sectors for illustration. The figure also shows how the ACF persists in time. This feature, obtained by the analysis of our data, is extremely striking and has not been reported in literature before. Further analysis is required, in future works, to fully understand this feature.

The study of the ACF is an empirical test of the efficiency of the BSE market for the period under consideration. The persistence of auto correlation we see above clearly indicates that the BSE is not an efficient market. For example Fig. 3 shows the significant consistent auto correlation in REALTY (lags 1, 2, 3, 7, 8, 9, 16, 17, 26), FMCG (lags 14, 23) and Sensex (lags 1, 8, 14, 17, 20, 23). Without high frequency data it is not possible to comment why these exact days lag are significant. However this broad analysis shows that during the period under consideration the BSE was not even weak efficient.

According to the Efficient Market Hypothesis (EMH), the stock prices fully reflect any changes in the information available to investors. For example, a market following random walk is consistent with the EMH. It has been shown [14] that mature stock markets are generally *weak* efficient. A departure from weak efficiency (i.e. deviation from random walk) may point towards possible market manipulation.

The auto correlation exhibited by the BSE sectors agrees with the findings in Ref. [15]. Those authors also show that auto correlation in returns might generate a momentum. Therefore, a BSE sector that outperformed other sectors in the past might continue to do so for some time interval. These features in the auto correlation may be crucial for portfolio management in Indian equity markets. Financial market volatility is central to the theory and practice of asset pricing, asset allocation, and risk management. Popular assumption is that volatilities and correlations are constant, but we have seen that they have significant variation over time. Therefore, the study of β can be useful for investor [16]. The β factor is defined as

$$\beta_i = \frac{\text{covariance}(i, Sensex)}{\text{variance}(Sensex)}. \tag{14}$$

A β of 1 indicates that the security's price will move with the market. A β of less than 1 means that the security will be less volatile than the market. A β of greater than 1 indicates that the security's price will be more volatile than the market. As an
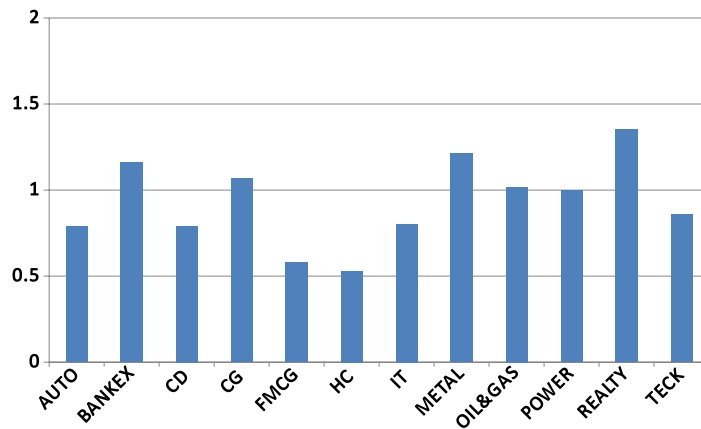
Fig. 4. β of all the sectors for April 2006–Mar 2014.

example, from Fig. 4 we can see that the β of the Bankex sector is 1.45 (45% more volatile than the market) while that of the HC sector is 0.49 (less volatile than the market). A systematic study of this parameter will be undertaken in a future work.

3. Cross-correlation among BSE sectors

In the last section we have shown that there is significant auto correlation in most sectors. Let us now try to see whether the movement of the indices in various sectors are also correlated i.e. whether there exists any cross correlation between the sectors. Some study of cross correlations of other markets have been carried out in different contexts [17] but to the best of our knowledge, there exists no studies of the correlations between sectors at least in the context of the Indian financial markets.

To understand the interactions among the sectors, it is useful to study the spectral properties of the correlation matrix of sectorial indices movements. The deviation of eigenvalues of the correlation matrix from those of a random matrix provide signals about the underlying interactions between various sectors. The largest eigenvalue is identified as representing the influence of the entire market, common for all sectors. The remaining large eigenvalues are associated with the different sectors, as indicated by the composition of their corresponding eigenvectors [18]. This is what we do in this subsequently in this section.

If the time series of returns of N sectors of length T are mutually uncorrelated, then the resulting correlation matrix is random and is known as Wishart matrix [19]. It is known that the empirical distribution of the eigenvalues of the Wishart matrix almost always converges to a probability distribution as $T \rightarrow \infty$ and $\frac{N}{T} \rightarrow a$ where a is a constant such that $0 \leq a < 1$. In that limit the distribution is continuous and supported on $\lambda \in [(1 - \sqrt{a})^2, (1 + \sqrt{a})^2]$ where $0 < a < 1$ [19]. This bound is known as the Random Matrix Theory (RMT) bound. Therefore, the eigenvalue λ of the Wishart matrix should lie between 0.84 and 1.17. We estimate the sample cross correlation matrix for our data set i.e. for $N = 13$ sectors for $T = 1990$ days and $\frac{N}{T} \rightarrow a = 0.006533$ (see Fig. 5). The reduced number of Principal Components (PC) of the cross correlation matrix that can explain most of the total variance is given in terms of the eigenvectors $\mathbf{u}$ of the cross correlation matrix as

$$\mathbf{C} = \sum_{i=1}^{N} \lambda_i \mathbf{u}_i \mathbf{u}_i^T. \tag{15}$$

We find the eigenvalues λ_i of the cross correlation matrix. The eigenvectors corresponding to these eigenvalues are the PCs of the cross correlation matrix. These eigenvectors can be expanded in a basis given by the 13 sectors we are considering. All the eigenvalues and the expansion of the PCs in our chosen basis is given in Fig. 6.

From Fig. 6 we can see significant deviation of the largest eigenvalue of the PC1 from the largest eigenvalue of RMT. The largest eigenvalue of the cross correlation matrix is 9.11. Also, from the first column (corresponding to PC1) of Fig. 6, the eigenvector of largest eigenvalue shows a relatively uniform composition, i.e. all sectors contribute to it and all elements having the same sign.

A very useful visualization of what we discussed above is the scree plot [20] as can be seen in Fig. 7. The fact that the PC1 is so large and that it affects all the sectors with the same ratio, we can say that the largest eigenvalue is associated with the collective response of the entire market to external informations [1,21], i.e. the largest eigenvalue is due to the existence of a market-induced correlation across all sectors. Since PC1 dominates to such a large extent it is difficult to observe the correlations between sectors.

From the investment point of view, it is interesting to note that the Tech and the IT sectors are highly correlated all the time. Hence, it would be better to club both these the sectors together for modeling and for portfolio diversification purposes.

Correlation Matrix

	AUTO	BANKEX	CD	CG	FMCG	HC	IT	METAL	OIL&GAS	POWER	REALTY	TECK	SENSEX
AUTO	1.00	0.72	0.64	0.74	0.59	0.67	0.53	0.74	0.70	0.75	0.65	0.65	0.82
BANKEX	0.72	1.00	0.61	0.79	0.56	0.62	0.53	0.72	0.73	0.79	0.73	0.66	0.89
CD	0.64	0.61	1.00	0.66	0.53	0.61	0.46	0.66	0.60	0.68	0.62	0.57	0.68
CG	0.74	0.79	0.66	1.00	0.56	0.65	0.53	0.75	0.74	0.88	0.72	0.67	0.86
FMCG	0.59	0.56	0.53	0.56	1.00	0.61	0.46	0.57	0.57	0.61	0.47	0.54	0.69
HC	0.67	0.62	0.61	0.65	0.61	1.00	0.53	0.66	0.65	0.70	0.57	0.64	0.73
IT	0.53	0.53	0.46	0.53	0.46	0.53	1.00	0.55	0.54	0.54	0.43	**0.93**	0.72
METAL	0.74	0.72	0.66	0.75	0.57	0.66	0.55	1.00	0.76	0.80	0.70	0.68	0.84
OIL&GAS	0.70	0.73	0.60	0.74	0.57	0.65	0.54	0.76	1.00	0.79	0.67	0.68	0.88
POWER	0.75	0.79	0.68	0.88	0.61	0.70	0.54	0.80	0.79	1.00	0.76	0.70	0.88
REALTY	0.65	0.73	0.62	0.72	0.47	0.57	0.43	0.70	0.67	0.76	1.00	0.58	0.74
TECK	0.65	0.66	0.57	0.67	0.54	0.64	**0.93**	0.68	0.68	0.70	0.58	1.00	0.85
SENSEX	0.82	0.89	0.68	0.86	0.69	0.73	0.72	0.84	0.88	0.88	0.74	0.85	1.00

Fig. 5. Cross-correlation matrix for April 2006–Mar 2014.

Eigenvectors

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13
Eigenvalues	9.11	0.91	0.62	0.46	0.37	0.32	0.31	0.29	0.24	0.21	0.11	0.05	0.01
AUTO	0.28	0.10	0.07	0.00	-0.14	-0.45	-0.61	0.45	0.07	0.32	-0.06	-0.02	-0.06
BANKEX	0.29	0.14	-0.19	0.26	0.16	0.09	-0.31	-0.21	**0.51**	-0.53	-0.10	-0.07	-0.27
CD	0.25	0.16	0.16	**-0.83**	0.30	-0.16	0.10	-0.17	0.22	-0.02	-0.01	0.01	0.03
CG	0.29	0.18	-0.17	0.08	0.04	-0.07	-0.18	-0.50	**-0.40**	0.18	**0.59**	-0.01	-0.11
FMCG	0.23	-0.02	0.77	0.31	0.44	0.15	0.05	0.08	-0.10	0.03	0.04	-0.02	-0.09
HC	0.26	-0.01	0.40	-0.14	**-0.78**	0.32	-0.05	-0.14	0.09	-0.11	0.06	-0.02	0.01
IT	0.23	**-0.71**	-0.12	-0.10	0.08	0.06	-0.05	0.02	-0.05	0.02	0.03	**-0.63**	0.01
METAL	0.29	0.11	-0.07	-0.03	-0.09	-0.28	0.29	0.39	**-0.44**	-0.61	0.08	-0.02	-0.06
OIL&GAS	0.29	0.07	-0.09	0.24	-0.14	-0.25	**0.63**	0.04	**0.41**	0.35	0.11	-0.10	-0.22
POWER	0.30	0.18	-0.09	0.08	-0.03	0.02	0.06	-0.30	-0.36	0.18	**-0.77**	-0.10	-0.01
REALTY	0.26	0.28	-0.31	-0.11	0.13	**0.69**	0.02	**0.45**	-0.03	0.20	0.10	-0.04	0.02
TECK	0.28	**-0.53**	-0.13	-0.05	0.04	0.07	0.01	0.01	-0.04	0.05	-0.07	**0.74**	-0.26
SENSEX	0.32	-0.06	-0.06	0.19	0.05	-0.08	0.03	-0.06	0.14	-0.05	0.05	0.17	**0.88**

Fig. 6. Eigenvalues and eigenvectors of correlation matrix for April 2006–Mar 2014.

The scree plot also gives some very useful information about periods of large fluctuations. During the time of large fluctuations we find that there is a is large correlation among most of the sectors. As a comparison consider Fig. 8 where we compare the cross correlation matrices of a period of large fluctuation (April 2008–March 2009) with a period of relatively small fluctuation (April 2012–Mar 2013). As can be seen from the figure, although there exists significant cross correlations at both the times, the magnitude is lesser in the later period. This indicates that periods of large fluctuations can be studied using models where the correlation strength becomes large. Since periods of large fluctuations may correspond to crashes
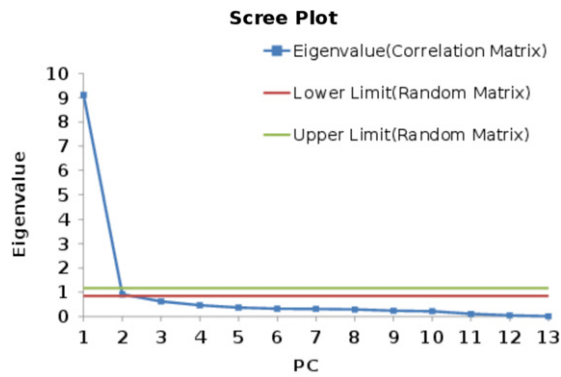
Scree Plot

Fig. 7. Eigenvalue and eigenvector matrix of correlation matrix for April 2006–Mar 2014 and scree plot for April 2006–Mar 2014.

Correlation Matrix 2008

	AUTO	BANKEX	CD	CG	FMCG	HC	IT	METAL	OIL&GAS	POWER	REALTY	TECK	SENSEX
AUTO	1.00	0.78	0.68	0.78	0.65	0.70	0.62	0.75	0.77	0.80	0.72	0.73	0.84
BANKEX	0.78	1.00	0.68	0.85	0.69	0.70	0.66	0.76	0.79	0.86	0.78	0.79	0.93
CD	0.68	0.68	1.00	0.74	0.59	0.67	0.62	0.69	0.64	0.75	0.70	0.71	0.74
CG	0.78	0.85	0.74	1.00	0.63	0.70	0.66	0.77	0.80	**0.92**	0.78	0.80	0.90
FMCG	0.65	0.69	0.59	0.63	1.00	0.64	0.55	0.59	0.59	0.68	0.58	0.64	0.72
HC	0.70	0.70	0.67	0.70	0.64	1.00	0.65	0.70	0.67	0.73	0.65	0.75	0.76
IT	0.62	0.66	0.62	0.66	0.55	0.65	1.00	0.66	0.64	0.67	0.56	**0.92**	0.79
METAL	0.75	0.76	0.69	0.77	0.59	0.70	0.66	1.00	0.82	0.81	0.72	0.79	0.87
OIL&GAS	0.77	0.79	0.64	0.80	0.59	0.67	0.64	0.82	1.00	0.84	0.76	0.79	0.92
POWER	0.80	0.86	0.75	**0.92**	0.68	0.73	0.67	0.81	0.84	1.00	0.80	0.83	0.93
REALTY	0.72	0.78	0.70	0.78	0.58	0.65	0.56	0.72	0.76	0.80	1.00	0.72	0.81
TECK	0.73	0.79	0.71	0.80	0.64	0.75	**0.92**	0.79	0.79	0.83	0.72	1.00	0.92
SENSEX	0.84	0.93	0.74	0.90	0.72	0.76	0.79	0.87	0.92	0.93	0.81	0.92	1.00

(a) April 2008–March 2009.

Correlation Matrix 2012

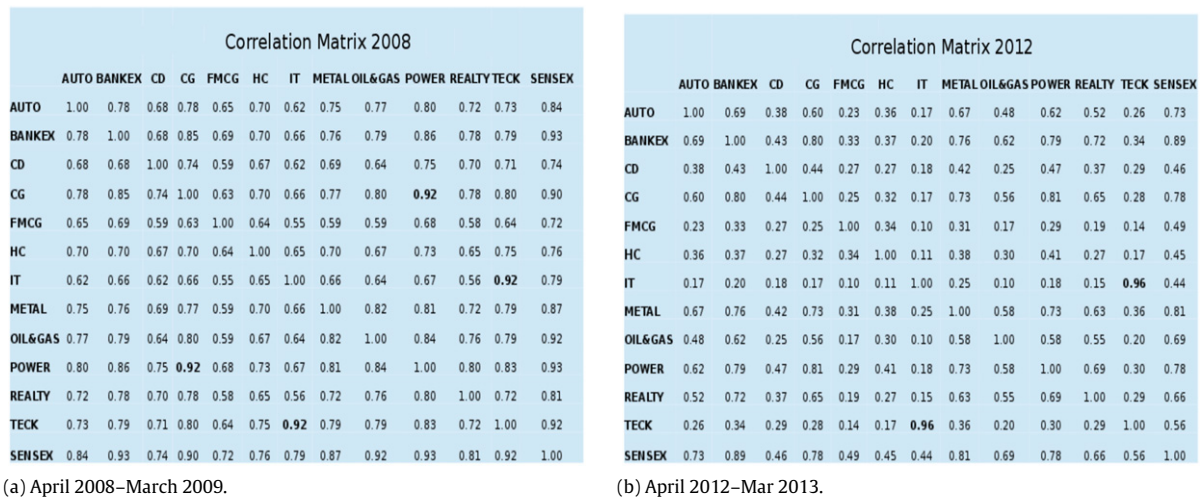|	AUTO	BANKEX	CD	CG	FMCG	HC	IT	METAL	OIL&GAS	POWER	REALTY	TECK	SENSEX
AUTO	1.00	0.69	0.38	0.60	0.23	0.36	0.17	0.67	0.48	0.62	0.52	0.26	0.73
BANKEX	0.69	1.00	0.43	0.80	0.33	0.37	0.20	0.76	0.62	0.79	0.72	0.34	0.89
CD	0.38	0.43	1.00	0.44	0.27	0.27	0.18	0.42	0.25	0.47	0.37	0.29	0.46
CG	0.60	0.80	0.44	1.00	0.25	0.32	0.17	0.73	0.56	0.81	0.65	0.28	0.78
FMCG	0.23	0.33	0.27	0.25	1.00	0.34	0.10	0.31	0.17	0.29	0.19	0.14	0.49
HC	0.36	0.37	0.27	0.32	0.34	1.00	0.11	0.38	0.30	0.41	0.27	0.17	0.45
IT	0.17	0.20	0.18	0.17	0.10	0.11	1.00	0.25	0.10	0.18	0.15	**0.96**	0.44
METAL	0.67	0.76	0.42	0.73	0.31	0.38	0.25	1.00	0.58	0.73	0.63	0.36	0.81
OIL&GAS	0.48	0.62	0.25	0.56	0.17	0.30	0.10	0.58	1.00	0.58	0.55	0.20	0.69
POWER	0.62	0.79	0.47	0.81	0.29	0.41	0.18	0.73	0.58	1.00	0.69	0.30	0.78
REALTY	0.52	0.72	0.37	0.65	0.19	0.27	0.15	0.63	0.55	0.69	1.00	0.29	0.66
TECK	0.26	0.34	0.29	0.28	0.14	0.17	**0.96**	0.36	0.20	0.30	0.29	1.00	0.56
SENSEX	0.73	0.89	0.46	0.78	0.49	0.45	0.44	0.81	0.69	0.78	0.66	0.56	1.00

(b) April 2012–Mar 2013.

Fig. 8. Comparison of correlation matrices.

in the stock market, a systematic study of the cross correlation matrices of these periods will provide valuable insights into understanding and modeling crashes.

A more efficient way of analyzing this would be by doing the Principle Component analysis we had performed previously in this section. Again, scree plots provide a more efficient and rigorous demonstration of the increase in correlations during periods of crisis. As can be see in Fig. 9, the PC1 when the entire market is experiencing large fluctuations is 9.91, while it comes down to 6.72 during period of relative calm. We can actually *zoom in* to the actual time of the crash (Jan 2008) in Fig. 10 using the quarterly and monthly data and see that the PC1 is actually higher (11.32) during that time. This can provide a efficient and novel way of analyzing crashes of the stock market.

4. Conclusions

In this paper, we have carried out a model independent analysis of the BSE for a period of 1990 days. This time frame contains periods of both small and large fluctuations and thus provides a good sample to understand and study the generic behavior of the stock market. Also the number of days chosen was large to avoid small sample size errors. Instead of studying the movement of individual stock returns as is usually done, we study the movement and behavior of groups of stocks, the grouping being done in terms of sectors. We look at 12 sectors of stocks and use the whole Sensex as the benchmark. The auto correlations in the return data captures how the stocks within the individual sectors interact among themselves while the cross correlations look at how the sectors affect each other.

We found the presence of significant auto correlations in all the sectors clearly demonstrating that the movement of the stock prices cannot be modeled via random walk. While this is usually a accepted feature of stock market model, our analysis of the departure from normality is rigorous. It is not just based on the non zero skewness and kurtosis but also on D'Agostino–Pearson omnibus test. This comprehensively shows the existence of auto correlations. From an investors point of view, this means that the only mean–variance–skewness–kurtosis based methods of portfolio optimization will be useful.
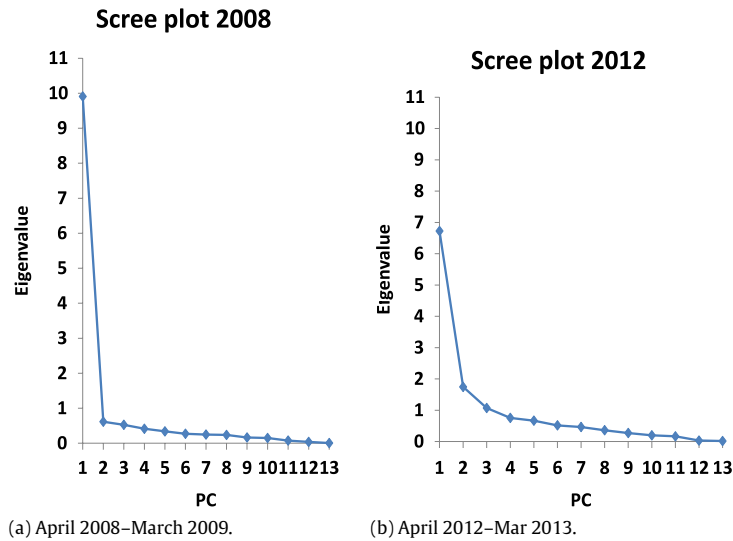
Scree plot 2008

Scree plot 2012

(a) April 2008–March 2009.

(b) April 2012–Mar 2013.

Fig. 9. Comparison of scree plots.

Scree plot Jan-Mar 2008
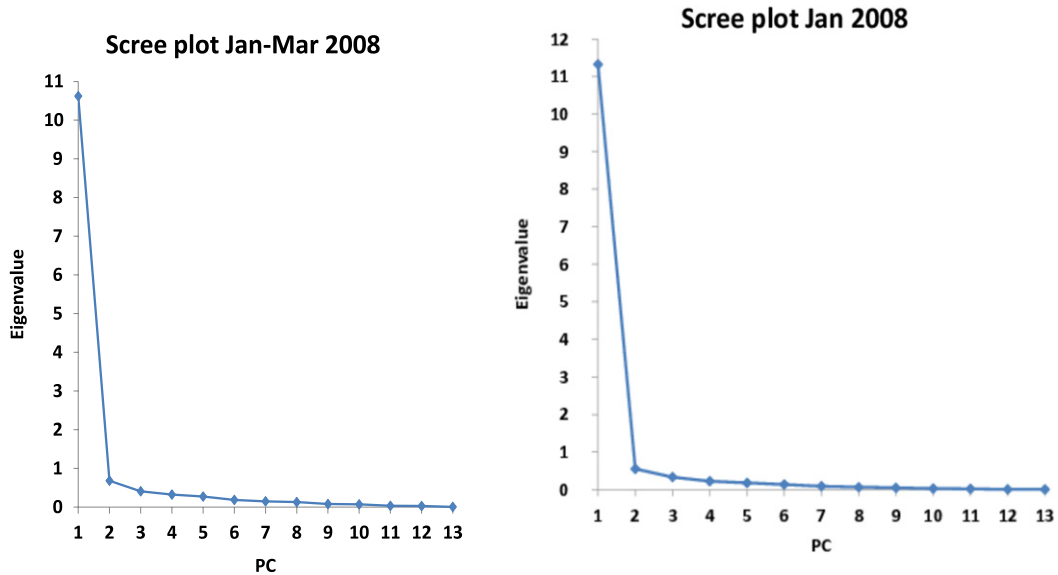
Scree plot Jan 2008

Fig. 10. Scree plot for April Jan–Mar 2008 and for Jan 2008.

A more interesting feature which we find in the study of auto correlations is that they persist over time. The ACF is significant for all sectors at lag one and there are certain sectors where this auto correlation persists at higher lags. This indicates that the BSE has significant departure from efficient market and EMH cannot be used to model the stock price movement in BSE. This is a very interesting property of the stock market which has to be accounted for in the future models. For financial markets to be meaningful and useful to the economy, they must be at least weakly efficient. Some of the reasons why BSE is not efficient may be (i) weak disclosure procedure (ii) poor quality and quantity of company' disclosure (iii) almost no public awareness about securities (iv) no transparent regulation, supervision and administrative rule. This feature of the BSE should be of great interest not just to the investors but also policy makers and market regulators. Further analysis of this will be done in a future work.

We also study the relative volatility of the sectors compared to the whole market, measured in terms of β. This parameter, as we point out, should have a significant role in making investment decisions. How to use this parameter in building physics models of financial markets is a direction of future work.

The cross correlation was studied by doing the Principle Component analysis of the correlation matrix. Our findings show that there exist a very large cross correlation but that correlation is due to some external force which drives the market as a whole. The effect of sectors on each other is smaller but not insignificant and will be the focus of a future work. A very

important feature following from our analysis is that the value of PC1 increases during periods of large fluctuation of the market. This can have far reaching application in studying and predicting crashes of financial markets.

References

[1] R.N. Mantegna, H.E. Stanley, Introduction to Econophysics, Cambridge University Press, Cambridge, 1999;
    Sitabhra Sinha, Arnab Chatterjee, Anirban Chakraborti, Bikas K. Chakrabarti, Econophysics: An Introduction, Wiley-VCH, Weinheim, 2010.
[2] T. Lux, The stable paretian hypothesis and the frequency of large returns: an examination of major German stocks, Appl. Financ. Econ. 6 (1996) 463–475.
[3] Dennis P. Quinn, Hans-Joachim Voth, A century of global equity market correlations, Amer. Econ. Rev.: Pap. Proc. 98 (2) (2008) 535–540;
    Fabrizio Lillo, Rasarion N. Mantegna, Ensemble properties of securities traded in the NASDAQ market, Physica A 299 (1–2) (2001) 161–167.
[4] S. Sinha, R.K. Pan, The power (law) of Indian markets: Analyzing NSE and BSE trading statistics, in: A. Chatterjee, B.K. Chakrabarti (Eds.), Econophysics of Stock and Other Markets, Springer, 2006, pp. 24–34;
    S. Sinha, R.K. Pan, Uncovering the internal structure of the Indian financial market: Large cross-correlation behavior in the NSE, in: A. Chatterjee, B.K. Chakrabarti (Eds.), Econophysics of Markets and Business Networks, Springer, 2007, pp. 3–19.
[5] http://www.nyse.com/.
[6] http://www.bseindia.com/.
[7] Ralph B. D'Agostino, Albert Belanger, Ralph B. D'Agostino Jr., A suggestion for using powerful and informative tests of normality, Amer. Statist. 44 (4) (1990) 316–321. JSTOR 2684359;
    D.N. Joanes, C.A. Gill, Comparing measures of sample skewness and kurtosis, Statistician 47 (1) (1998) 183–189;
    R. D'Agostino, E.S. Pearson, Tests for departure from normality, Biometrika 60 (3) (1973) 613–622.
[8] V. Plerou, P. Gopikrishnan, H.E. Stanley, Two-phase behavior of financial markets, Nature 421 (2003) 130.
[9] Eugene F. Fama, Random walks in stock market prices, Financ. Anal. J. 21 (5) (1965) 55–59.
[10] H. Markowitz, Portfolio selection, J. Finance (7) (1952) 77–91.
[11] K.K. Lai, L. Yu, S. Wang, Mean–variance–skewness–kurtosis based portfolio optimization, in: Proceeding of the First International Multi-Symposiums on Computer and Computational Sciences, IMSCCS 06, Vol. 2, 2006, pp. 292–297;
    P. Jena, T.K. Roy, S.K. Majumdar, Multi-objective mean–variance–skewness model for portfolio optimization, Adv. Model. Optim. 9 (1) (2007) 181–193.
[12] G.E.P. Box, G.M. Jenkins, G.C. Reinsel, Time Series Analysis: Forecasting Control, fourth ed., Wiley, 2008.
[13] M.S. Bartlett, On the theoretical specification and sampling properties of autocorrelated time series, J. R. Stat. Soc. Ser. B 8 (1946) 27–41.
[14] E. Fama, The behavior of stock market prices, J. Bus. 38 (1965) 34–105;
    E. Fama, Efficient capital markets: A review of theory and empirical work, J. Finance 25 (2) (1970) 383–417.
[15] A.W. Lo, A.C. MacKinlay, A Non-Random Walk Down Wall Street, Princeton University Press, 1999, 14–16 and Chapter 2.
[16] Myron Scholes, Joseph Williams, Estimating betas from nonsynchronous data, J. Financ. Econ. 5 (3) (1977) 309–327;
    Seth Klarman, Joseph Williams, Beta, J. Financ. Econ. 5 (3) (1991).
[17] R.K. Pan, S. Sinha, Collective behavior of stock price movements in an emerging market, Phys. Rev. E (76) (2007) 046116. 1–9.
[18] V. Plerou, P. Gopikrishnan, L.A.N. Amaral, M. Meyer, H.E. Stanley, Scaling of the distribution of price fluctuations of individual companies, Phys. Rev. E (60) (1999) 6519–6529;
    P. Gopikrishnan, V. Plerou, L.A.N. Amaral, M. Meyer, H.E. Stanley, Scaling of the distribution of fluctuations of financial market indices, Phys. Rev. E 60 (1990) 5305–5316;
    M. Sarma, Characterisation of the Tail Behaviour of Financial Returns: Studies from India, EURANDOM Report 2005-003, 2005, (http://www.eurandom.tue.nl/reports/2005/003MSreport.pdf);
    K. Matia, M. Pal, H. Salunkay, H.E. Stanley, Scale-dependent price fluctuations for the Indian stock market, Europhys. Lett. 66 (2004) 909–914.
[19] D. Johnson, Some limit theorems for the eigenvalues of a sample covariance matrix, J. Multivariate Anal. (12) (1982) 1–38;
    V.A. Marchenko, L.A. Pastur, The distribution of eigenvalues in certain sets of random matrices, Mat. Sb. (72) (1967) 507–536.
[20] R.B. Cattell, The scree test for the number of factors, Multivariate Behav. Res. 1 (1966) 245–276.
[21] L. Laloux, P. Cizeau, J.P. Bouchaud, M. Potters, Noise dressing of financial correlation matrices, Phys. Rev. Lett. (83) (1999) 1467;
    V. Plerou, P. Gopikrishnan, B. Rosenow, L.A. Nunes Amaral, H.E. Stanley, Universal and nonuniversal properties of cross correlations in financial time series, Phys. Rev. Lett. (83) (1999) 1471.